On this page ⌄

Invoke a Job

Asynchronous Endpoints

Useful for long-running tasks. Submit a job and use the returned Job ID to check the status later.

Example:

```
{
  "id": "eaebd6e7-6a92-4bb8-a911-f996ac5ea99d",
  "status": "IN_QUEUE"
}
```

Synchronous Endpoints

Ideal for short tasks where immediate results are preferred.

Example:

```
{
  "delayTime": 824,
  "executionTime": 3391,
  "id": "sync-79164ff4-d212-44bc-9fe3-389e199a5c15",
  "output": [
    {
      "image": "https://image.url",
      "seed": 46578
    }
  ],
  "status": "COMPLETED"
}
```

Ask AI

Health Endpoint

The `/health` endpoint provides information on worker numbers and additional statistics for an endpoint.

Example response:

```json
{
  "jobs": {
    "completed": 1,
    "failed": 5,
    "inProgress": 0,
    "inQueue": 2,
    "retried": 0
  },
  "workers": {
    "idle": 0,
    "running": 0
  }
}
```

Purge Queue Endpoint

The `/purge-queue` endpoint clears all jobs in the queue but does not affect jobs in progress.

Example response:

```json
{
  "removed": 0,
  "status": "completed"
}
```

Starting and Monitoring Jobs

Start a Job

```bash
curl -X POST https://api.runpod.ai/v2/{endpoint}/run \
    -H 'Content-Type: application/json' \
    -H 'Authorization: Bearer {API key}' \
    -d '{"input": {"prompt": "Your prompt"}}'
```

Ask AI

Check Job Status

```
curl https://api.runpod.ai/v2/{endpoint}/status/{job_id} \
    -H 'Authorization: Bearer {API key}'
```

Rate Limits

- `/run`: 1000 requests every 10 seconds.
- `/runsync`: 2000 requests every 10 seconds.

> ⓘ **NOTE**
>
> Retrieve results within 30 minutes for privacy protection.

✏️ Edit this page

Docs

Overview

Tutorials

AI APIs

Community

Discord ⤴

Contact us ⤴

Ask AI

More

Blog 🔗

GitHub 🔗

Ask AI